BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e36207
doi: 10.3897/biss.3.36207

OPEN ACCESS

Conference Abstract

Linking Biodiversity Data Using Evolutionary History

Emily Jane McTavish ‡

‡ University of California, Merced, United States of America

Corresponding author: Emily Jane McTavish (ejmctavish@ucmerced.edu)

Received: 14 May 2019 | Published: 21 Jun 2019

Citation: McTavish E (2019) Linking Biodiversity Data Using Evolutionary History. Biodiversity Information Science and Standards 3: e36207. https://doi.org/10.3897/biss.3.36207

Abstract

All life on earth is linked by a shared evolutionary history. Even before Darwin developed the theory of evolution, Linnaeus categorized types of organisms based on their shared traits. We now know these traits derived from these species' shared ancestry. This evolutionary history provides a natural framework to harness the enormous quantities of biological data being generated today.

The Open Tree of Life project is a collaboration developing tools to curate and share evolutionary estimates (phylogenies) covering the entire tree of life (Hinchliff et al. 2015, McTavish et al. 2017). The tree is viewable at https://tree.opentreeoflife.org, and the data is all freely available online. The taxon identifiers used in the Open Tree unified taxonomy (Rees and Cranston 2017) are mapped to identifiers across biological informatics databases, including the Global Biodiversity Information Facility (GBIF), NCBI, and others. Linking these identifiers allows researchers to easily unify data from across these different resources (Fig. 1). Leveraging a unified evolutionary framework across the diversity of life provides new avenues for integrative wide scale research. Downstream tools, such as R packages developed by the R OpenSci foundation (rotl, rgbif) (Michonneau et al. 2016, Chamberlain 2017) and others tools (Revell 2012), make accessing and combining this information straightforward for students as well as researchers (e.g. https://mctavishlab.github.io/BIO144/labs/rotl-rgbif.html).

© McTavish E. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

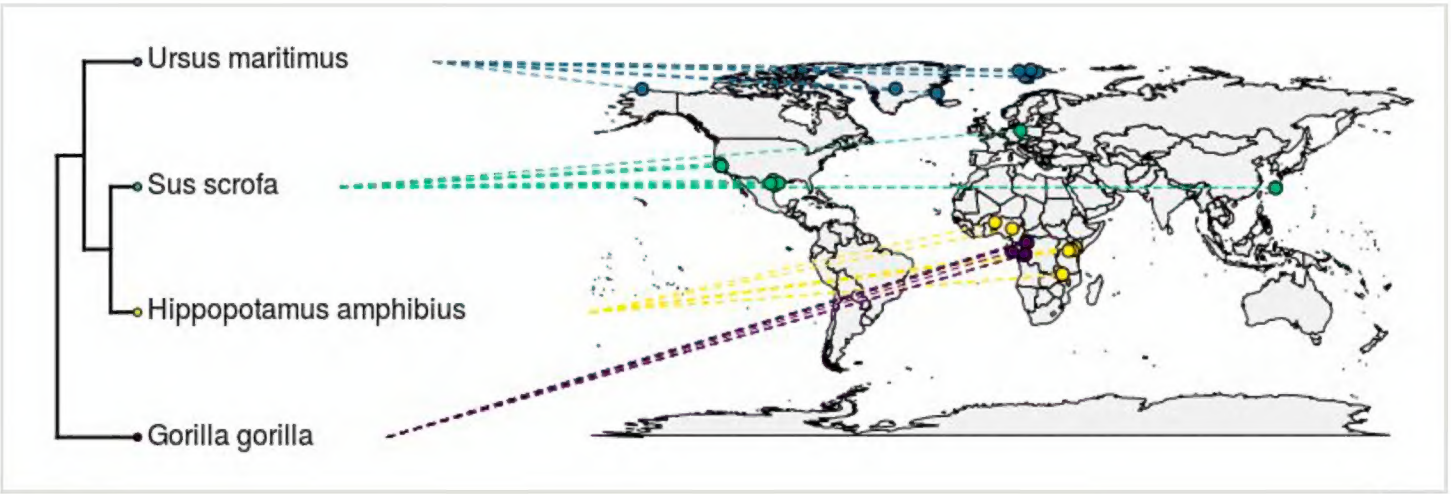

Figure 1.

Example linking phylogenetic relationships accessed from the Open Tree of Life with specimen location data from Global Biodiversity Information Facility.

For example, a recent publication by Santorelli et al. 2018 linked evolutionary information from Open Tree with species locality data gathered from a local field study as well as GBIF species location records to test a river-barrier hypothesis in the Amazon. By combining these data, the authors were able test a widely held biogeographic hypothesis across 1952 species in 14 taxonomic groups, and found that a river that had been postulated to drive endemism, was in fact not a barrier to gene flow. However, data provenance and taxonomic name reconciliation remain key hurdles to applying data from these large digital biodiversity and evolution community resources to answering biological questions. In the Amazonian river analysis, while they leveraged use of GBIF records as a secondary check on their species records, they relied on their an intensive local field study for their major conclusions, and preferred taxon specific phylogenetic resources over Open Tree where they were available (Santorelli et al. 2018). When Li et al. 2018 assessed large scale phylogenetic approaches, including Open Tree, for measuring community diversity, they found that synthesis phylogenies were less resolved than purpose-built phylogenies, but also found that these synthetic phylogenies were sufficient for community level phylogenetic diversity analyses. Nonetheless, data quality concerns have limited adoption of analyses data from centralized resources (McTavish et al. 2017). Taxonomic name recognition and reconciliation across databases also remains a hurdle for large scale analyses, despite several ongoing efforts to improve taxonomic interoperability and unify taxonomies, such at Catalogue of Life + (Bánki et al. 2018).

In order to support innovative science, large scale digital data resources need to facilitate data linkage between resources, and address researchers' data quality and provenance concerns. I will present the model that the Open Tree of Life is using to provide evolutionary data at the scale of the entire tree of life, while maintaining traceable provenance to the publications and taxonomies these evolutionary relationships are inferred from. I will discuss the hurdles to adoption of these large scale resources by researchers, as well as the opportunities for new research avenues provided by the connections between evolutionary inferences and biodiversity digital databases.

Keywords

evolution, phylogeny, taxonomy, interoperability

Presenting author

Emily Jane McTavish

Presented at

Biodiversity_Next 2019

Funding program

NSF Division Of Biological Infrastructure, Advances in Biological Informatics #1759846

Grant title

"Cultivating a sustainable Open Tree of Life"

References

- Bánki O, Döring M, Holleman A, Addink W (2018) Catalogue of Life Plus: innovating the CoL systems as a foundation for a clearinghouse for names and taxonomy. Biodiversity Information Science and Standards 2 https://doi.org/10.3897/biss.2.26922
- Chamberlain S (2017) rgbif: Interface to the Global 'Biodiversity' Information Facility API. 0.99. URL: https://CRAN.R-project.org/package=rgbif
- Hinchliff C, Smith S, Allman J, Burleigh JG, Chaudhary R, Coghill L, Crandall K, Deng J, Drew B, Gazis R, Gude K, Hibbett D, Katz L, Laughinghouse HD, McTavish EJ, Midford P, Owen C, Ree R, Rees J, Soltis D, Williams T, Cranston K (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proceedings of the National Academy of Sciences 112 (41): 12764-12769. https://doi.org/10.1073/pnas.1423041112
- Li D, Trotta L, Marx H, Allen J, Sun M, Soltis D, Soltis P, Guralnick R, Baiser B (2018) For comparing phylogenetic diversity among communities, go ahead and use synthesis phylogenies. bioRxiv https://doi.org/10.1101/370353
- McTavish EJ, Drew B, Redelings B, Cranston K (2017) How and Why to Build a Unified Tree of Life. BioEssays 39 (11). https://doi.org/10.1002/bies.201700114
- Michonneau F, Brown J, Winter D (2016) rotl: an R package to interact with the Open Tree of Life data. Methods in Ecology and Evolution n/a-n/a-n/a-n/a. https://doi.org/10.1111/2041-210X.12593

- Rees J, Cranston K (2017) Automated assembly of a reference taxonomy for phylogenetic data synthesis. Biodiversity Data Journal https://doi.org/10.3897/BDJ.5.e12581
- Revell L (2012) phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3 (2): 217-223. https://doi.org/10.1111/j.2041-210X.2011.00169.x
- Santorelli S, Magnusson W, Deus C (2018) Most species are not limited by an Amazonian river postulated to be a border between endemism areas. Scientific Reports 8 (1). https://doi.org/10.1038/s41598-018-20596-7